THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 60, No. 4, April 1981
Printed in U.S.A.

Subsampling of a DPCM Speech Channel to Provide Two "Self-Contained" Half-Rate Channels

By N. S. JAYANT

(Manuscript received November 12, 1980)

We consider the following channel-splitting problem: it is required to split a B-bits/s speech-code sequence into two "self-contained" B/2-bits/s components, either of which can be used to reproduce acceptable speech; also, if both components are available at a receiver, it must be possible to reproduce speech with full B-bits/s quality. We propose a solution where for interesting values of B, the speech quality resulting from half-rate receptions approximately equals that from conventional full-rate receptions at B/2 bits/s. In the proposed solution, 3.2-kHz speech is sampled at 12 kHz and coded using PCM *or differential* PCM. *The output sequence of codewords is split into odd- and even-word sequences. A full-rate receiver with access to both of the subchannels simply reconstitutes the output sequence prior to decoding, while a half-rate receiver with only the odd (or even) subchannel estimates the even (or odd) components by nearest-neighbor interpolation.*

1. INTRODUCTION

The channel-splitting problem described in the abstract is redefined in Fig. 1. The receiving end of a speech communication system is supposed to operate in either a full-rate or half-rate mode depending on whether it has available to it both or only one of the speech subchannels. Respective qualities of speech reproduction are denoted by $Q_F(B)$ and $Q_H(B/2)$. The nonavailability of one subchannel is a good model for certain types of transmission failure, examples of which are signal fading in mobile radio and speech segment losses in packet switching. With appropriate forms of diversity reception, the second subchannel will be available with probability close to unity when the first subchannel is not. The channel-splitting problem has been re-

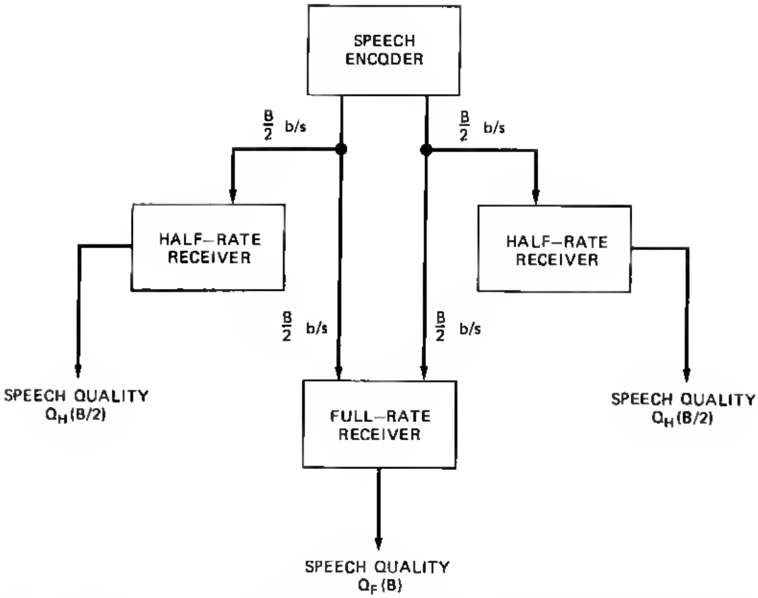


Fig. 1—Definition of the channel-splitting problem. When both of two half-rate components in the transmitted sequence are available at a receiver, conventional full-rate receptions result, with speech quality $Q_F(B)$. When only one of the components is present, a half-rate receiver recovers an approximation to the full-rate speech, with quality $Q_H(B/2)$.

cently analyzed for communication systems operating at rate-distortion limits with binary and Gaussian input signals.[1,2]

The nontrivial nature of the channel-splitting problem can be appreciated through the simple example of a uniform quantizer. By combining two appropriately staggered R-bit quantizers (one of them a midrise, the other a midtread), one can realize an $(R + 1)$-bit system, but not a $2R$-bit system. For full-rate speech quality corresponding to 8-bit quantization, component quantizers would each need 7-bit (not $8/2 = 4$ bit) resolution for the combination to yield 8-bit quality. Thus, if the subchannels in Fig. 1 were simply uniform quantizers, and if speech were sampled at 8 kHz, one would need two 56-kbit/s quantizer systems so that a full-rate combination with 64-kbit/s quality can be realized. By contrast, in the differential pulse code modulation (DPCM) system proposed in this paper, the component receivers that combine to give 64-kbit/s quality are indeed half-bit rate, 32-kbit/s systems. Moreover, with an illustrative sentence-length speech input, the half-rate quality $Q_H(B/2)$ will be shown to exceed the full-rate quality $Q_F(B/2)$ of a conventional DPCM system operating at $B/2$ bit/s, for interesting values of B. The quality Q_F (32 kb/s) is quite acceptable

for speech communications, although somewhat short of toll quality. Note once again that two uniform quantizers operating at 32-kbit/s (8 kHz × 4 bits) each can only give, in combination, 40-kbit/s (8 kHz × 5 bits) quality, and not the desired 64-kbit/s quality. Similar arguments apply to lower bit rates as well.

The system of Fig. 1 is a special case of a communication scenario that can be generalized to include more than two subchannels, and/or subchannels that are non-equal-rate. The system of Fig. 1 can also be regarded as a special, symmetrical case of embedded coding,[3,4] with a hierarchy consisting of two equally significant subcodes, viz., the half-rate sequances.

II. SUBSAMPLING AND INTERPOLATION

The utility of subsampling and interpolation has been demonstrated recently in the context of speech packet losses;[5] speech-encoder outputs are partitioned into odd-sample and even-sample systems which are transmitted as separate packets. In the event of a lost odd (or even) packet, the lost samples are estimated using nearest-neighbor interpolations involving available even (or odd) samples. With the usual assumption of 3.2-kHz speech and 8-kHz sampling, the 1:2 subsampling (at 4 kHz) implies serious aliasing effects, but these errors are mitigated by an adaptive interpolation procedure where nearest-neighbor-weighting coefficients are varied to follow speech statistics, as reflected by appropriate extra information in packet headers.[5] The system realizes dramatic improvements with packet loss probabilities up to about 10 percent; but as the component message loss probability approaches 100 percent, as in the channel-splitting problem, residual aliasing effects are quite unacceptable, even with adaptive interpolation.

The above observation has led us to the notion of 12-kHz sampling for the problem of Fig. 1. With 1:2 subsampling, the half-sampling rate will now be 6-kHz, which turns out to be just adequate for the 3.2-kHz speech inputs in telephony. We also considered 16-kHz sampling, but this is less preferable from the point of view of quantization noise. In 64-kbit/s decoding for example, 4-bit quantization of 16-kHz speech produces more quantization noise than the 5-bit quantizer that is possible with 12-kHz speech.

A second advantage of 12-kHz sampling is that it permits nonadaptive interpolations; adaptive interpolations yielded near-zero gains with 12-kHz speech. If 8-kHz is subsampled, it cannot be adequately reconstituted by nonadaptive interpolation even if the interpolation is invoked with a probability much less than 100 percent.[5]

Waveform reconstructions in the half-bit-rate (half-sample rate)

systems of this paper are described by

$$\hat{u}(n) = A_1 \cdot u(n-1) + A_2 \cdot u(n+1),$$
$$A_1 = A_2 = 0.5. \tag{1}$$

The samples $u(r)$ will be quantized DPCM prediction error samples in general; and more simply, they will be quantized speech samples in the special case of nonpredictive, or nondifferential PCM.

III. THE DPCM CODEC

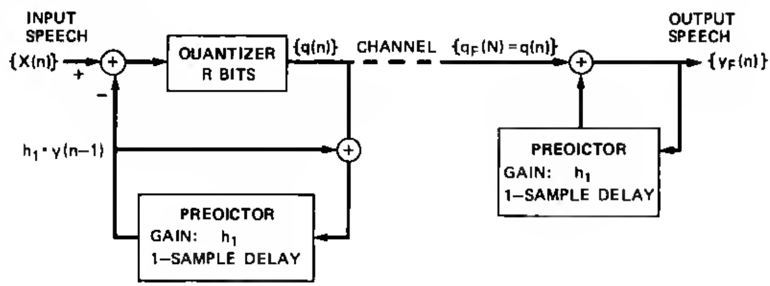
Figure 2 shows block diagrams of full-rate and half-rate DPCM systems with fixed first-order predictors. In each case, decoding is defined by

$$y(n)\,|_{\mathrm{DPCM}} = h_1 \cdot y(n-1) + q(n), \tag{2}$$

where $q(n)$ is the quantized prediction error signal, and h_1 is a first-order predictor. In the special case of nondifferential PCM ($h_1 = 0$), $q(n)$ is simply the quantized speech output:

$$y(n)\,|_{\mathrm{PCM}} = q(n). \tag{3}$$

Subscripts H and F in Fig. 2 distinguish balf- and full-rate versions of $y(n)$ and $q(n)$.



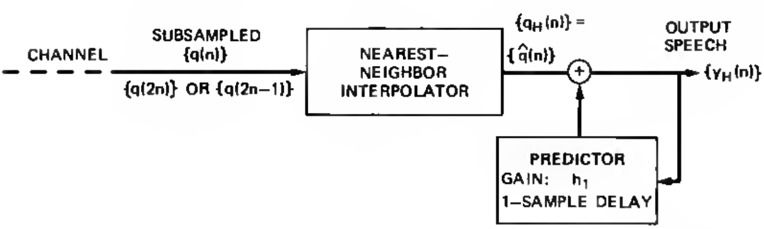


Fig. 2—DPCM block diagrams. (a) Conventional full-rate codec. (b) Decoder portion of half-rate receiver.

In contrast with well-known adaptive DPCM systems where the quantizer step size is adapted for every sample, the present paper assumes a system where the step size is adapted once for each block (of several ms duration), and held fixed for the duration of the block.[6] A periodically updated, rather than instantaneously adaptive, quantizer is used in anticipation of interpolation procedures, which are known from recent experience to be unreliable when the step size exhibits sample-to-sample fluctuation.[7]

The periodically adaptive quantizers are defined by a block-specific step size Δ that is proportional to the root-mean-square value of a generalized first-difference in the form

$$\Delta\,|_{AQF} = K_1(R)\cdot[x(n) - h_1\cdot x(n-1)]_{\mathrm{rms}}, \tag{4a}$$

$$\Delta\,|_{AQB} = K_1(R)\cdot[y(n) - h_1\cdot y(n-1)]_{\mathrm{rms}}, \tag{4b}$$

with maximum and minimum constraints

$$\Delta_{\max} = 128\cdot\Delta_{\min} = K_2(R)\cdot|x(n)|_{\max}, \tag{4c}$$

where $[K_1(R), K_2(R)]$ are bit-rate-specific constraints with suggested values of [0.25, 0.03], [0.33, 0.06], [0.58, 0.12] and [1.0, 0.18] for 5, 4, 3 and 2-bit quantizers respectively. The subscripts AQF (4a) and AQB (4b) refer to forward-adaptive and backward-adaptive procedures; respective rms values in (4a) and (4b) are evaluated over the duration of a speech block to be coded in AQF, and over the duration of the most recent decoded speech block in AQB. The AQB procedure is less effective because of speech nonstationarity as well as the effect of quantization noise that is present in the $y(n)$ sequence used in (4b). However, step-size information in an AQB system need not be separately transmitted to a receiver; it is inherently available in the decoded $y(n)$ sequence. AQF procedures, by constrast, require the explicit transmission of step-size information (typically, about 5 bits worth, per block of 16 ms). In our experiments, quality losses in AQB were more noticeable in full-bit-rate speech than in half-bit-rate speech; and in each case the losses were of a second order of importance. With this in mind we have elected to cite only AQF results in section IV; these results can be regarded as upper bounds as far as quantizer performance is concerned.

In the context of 1:2 subsampling, the AQB procedure of (4b) cannot be implemented as such unless $h_1 = 0$ (PCM). However, step sizes obtained by setting h_1 to zero in (4) have been found to have fairly small effects on DPCM performance. Differences between PCM-optimal and DPCM-optimal step size are less significant than differences among step sizes of different speech blocks. Moreover, the suboptimality of PCM-matched step size becomes less significant as $h_1 \to 0$. The next

section shows that half-bit-rate DPCM favor values of h_1 that are indeed much smaller than those appropriate for conventional full-rate DPCM.

IV. RESULTS AND CONCLUSIONS

Figures 3 through 6 illustrate the performance of the interpolation procedure for the example of a 5-bit encoder. The waveform segments refer to two 20-ms blocks from a 3.2-kHz bandlimited, female utterance "The chairman cast three votes" sampled at 12 kHz.

Figure 3 shows full-rate and interpolated q waveforms ($h_1 = 0.5$) for the two segments; it demonstrates that the nonadaptive interpolator (1) is reasonably adequate even for the fast-varying unvoiced example. This is confirmed in Fig. 4 which shows corresponding full-rate and half-rate y waveforms. Notice that the half-bit-rate output provides a much better reproduction of the voiced segment, but is nevertheless reasonably effective in unvoiced speech reproduction. Perceptually,

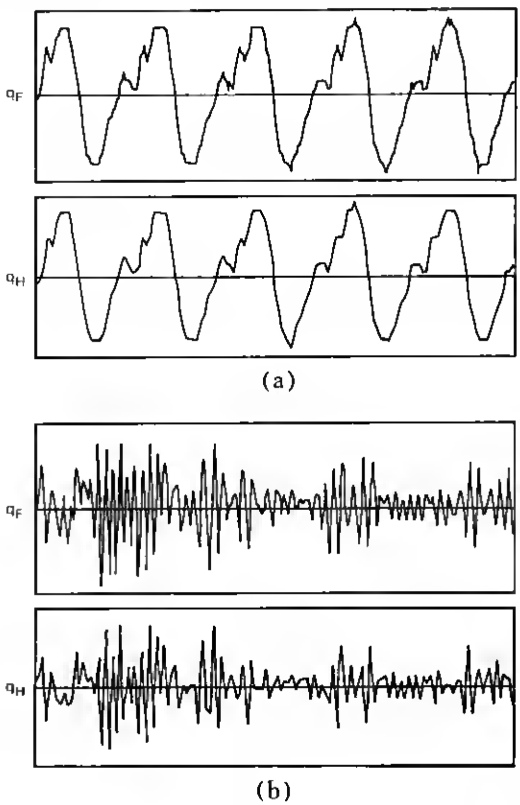


Fig. 3—Full-rate and half-rate (interpolated) sequences q_F and q_H (R = 60 kb/s, h_1 = 0.5) for (a) voiced speech segment and (b) unvoiced speech segment. The speech segments are 16 ms long.

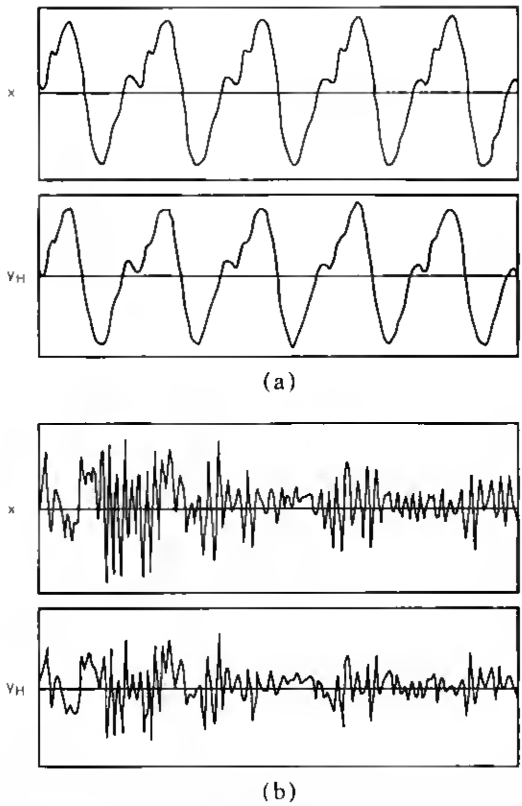


Fig. 4—Original and half-rate-decoded speech output $y(n)$ ($R = 60$ kb/s, $h_1 = 0.5$) for (a) voiced speech segment and (b) unvoiced speech segment. The speech segments are 16 ms long.

the waveform degradations in the half-bit coder are indeed fairly subtle for $R = 5$.

The significance of $h_1 = 0.5$ is demonstrated in Fig. 5, which sbows DPCM performance as a function of predictor coefficient value. The objective quality measure used is the segmental signal-to-noise ratio SEGSNR defined as the average of 10 log s/n (dB)-values measured over the totality of 20-ms blocks in the input. Notice that maximization of full-rate and half-rate quality call for $h_1 = 0.9$ and $h_1 = 0.5$, respectively; and notice also that these are not very sharp maxima, suggesting flexibility for practical implementations. The special, simple case of $h_1 = 0$ (PCM) leads to a noticeable quality degradation only for the full-rate system.

Figure 6 depicts the performance of full-rate and half-rate DPCM receivers as a function of encoder bit rate. The performance curves (i) and (ii) are for h_1 values that maximize half-rate speech quality (h_1

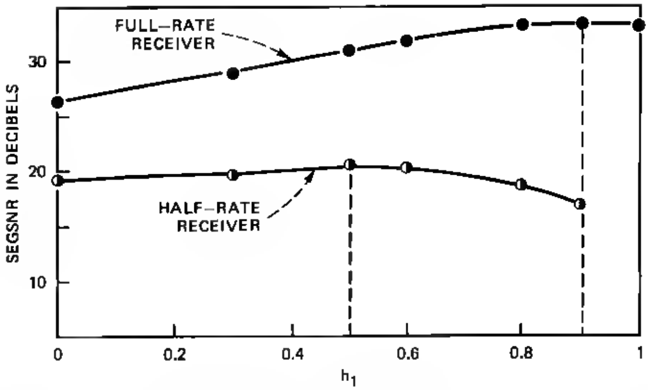


Fig. 5—Segmental SNR versus predictor coefficient h_1 (R = 60 kb/s). Values that maximize full-rate and half-rate quality are respectively 0.9 and 0.5.

= 0.5 for 5,4 bits/sample, h_1 = 0.3 for 3,2 bits/sample). The full-rate characteristic shows the expected 6-dB-per-bit behavior, while the half-bit-rate characteristic falls more gradually with decreasing R. Both characteristics tend to the expected 0-dB limit for no transmissions ($R \to 0$). The square dots in Fig. 6 represent the performance of a full-rate receiver in a system designed to maximize full-rate speech quality (h_1 = 0.8 to 0.9).

An important observation from Fig. 6 is that for encoder bit rates in the important range of 2 to 5 bits/sample,

$$Q_H(R/2) \approx Q_F(R/2);\ R < 5 \text{ bits/sample.} \tag{5}$$

This suggests that the half-bit-rate qualities in the subsample-inter-

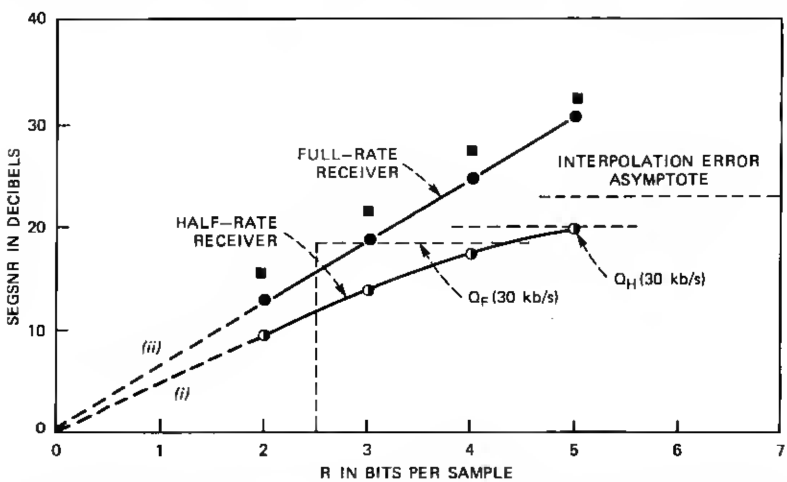


Fig. 6—Segmental SNR versus total transmitted bits/sample R. Curves (i) and (ii) refer to half-rate and full-rate codecs designed to maximize half-rate quality. The square dots represent a full-rate codec designed to maximize full-rate quality.

polate system are extremely good results, considering the crucial constraint that the half-bit-rate systems combine trivially to yield the full-rate performance $Q_F(R)$. The approximate equality in (5) is borne out very well in perceptual assessments of Q_H and Q_F. In contrast to (5), analytical results in Refs. 1 and 2 are quite pessimistic. This difference in conclusions is related to the fact that these analytical results apply at the rate-distortion limit, while the bit rates in this paper are nowhere close to the rate-distortion limit for speech. In fact, our bit rates are high enough that there is sufficient redundancy left in the coder output to permit subsampling and high-quality interpolations.

The relative performance of the half-bit-rate receiver diminishes with increasing bit rate. Clearly, as $R \rightarrow \infty$, the quantization noise contributions to Q vanish, $Q_F \rightarrow \infty$, and Q_H tends to a finite asymptotic value that shows the effect of nearest-neighbor interpolation noise. Results elsewhere[5] can be used to show that this asymptotic value, for a first-order Markov signal example, is approximately given by the expected value of $10 \log [(1 + R_{xx}^2(1))/(1 - R_{xx}^2(1)] \; dB$, where $R_{xx}(1)$ is a block-specific adjacent sample correlation in the speech input $x(n)$.

Finally it would be appropriate to calibrate the Q_F and Q_H values in Fig. 6 with well-known definitions of toll-quality and communications quality (near perfect intelligibility with noticeable but not obstrusive degradation).[8] Although simulations and conclusions have centered on a single earlier-cited test input, it appears that full-bit-rate DPCM realizes toll-quality for $R = 5$ and 4 bits/sample; and communications quality for $R = 3$ and 2 bits/sample. The half rate DPCM receptions in the proposed system approach toll quality with $R = 5$ bits/sample and maintain good communications quality at $R = 4$ and 3 bits/sample.

REFERENCES

1. J. K. Wolf, A. D. Wyner, and J. Ziv, "Source Coding for Multiple Descriptions," B.S.T.J., *59* (October 1980), pp. 1417–26.
2. L. Ozarow, "On a Source Coding Problem With Two Channels and Three Receivers," B.S.T.J., *59* (December 1980), pp. 1909–21.
3. D. J. Goodman, "Embedded DPCM for Variable Bit Rate Transmission," IEEE Trans. Commun., *COM-28*, No. 7 (July 1980), pp. 1040–46.
4. T. Bially, B. Gold, and S. Seneff, "A Technique for Adaptive Voice Flow Control in Integrated Packet Networks," IEEE Trans. Communications, *COM-28*, No. 3 (March 1980) pp. 325–33.
5. N. S. Jayant and S. W. Christensen, "Effects of Packet Losses in Waveform-Coded Speech and Improvements due to an Odd-Even Sample Interpolation Procedure," IEEE Trans. Commun., *COM-29* (February 1981).
6. N. S. Jayant, "Step-Size Transmitting Differential Coders for Mobile Telephony," B.S.T.J., *54* (November 1975), pp. 1557–81.
7. N. S. Jayant et al., "On Soft Decision Demodulation for PCM and DPCM Encoded Speech," IEEE Trans. Commun., *COM-28*, No. 3 (March 1980), pp. 334–44.
8. J. L. Flanagan et al., "Speech Coding," IEEE Trans. Commun., *COM-27* (April 1979), pp. 710–37.